

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平7-13905

(43) 公開日 平成7年(1995)1月17日

(51) Int.Cl.⁶
G 0 6 F 13/12

識別記号 庁内整理番号
3 1 0 C 8133-5B

F I

技術表示箇所

審査請求 未請求 請求項の数27 O L (全 25 頁)

(21) 出願番号 特願平5-151786

(22) 出願日 平成5年(1993)6月23日

(71) 出願人 000005108

株式会社日立製作所

東京都千代田区神田駿河台四丁目6番地

(72) 発明者 山神 憲司

神奈川県川崎市麻生区王禅寺1099番地 株

式会社日立製作所システム開発研究所内

(72) 発明者 渡辺 治明

神奈川県川崎市麻生区王禅寺1099番地 株

式会社日立製作所システム開発研究所内

(72) 発明者 横畑 静生

神奈川県小田原市国府津2880番地 株式会

社日立製作所ストレージシステム事業部内

(74) 代理人 弁理士 小川 勝男

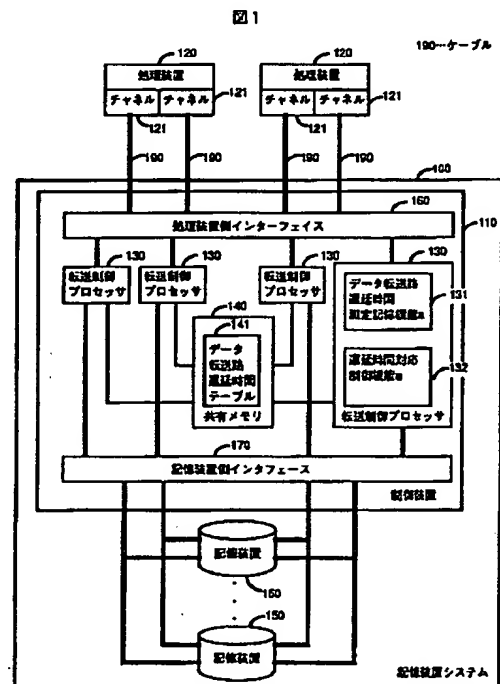
最終頁に続く

(54) 【発明の名称】 記憶装置システム及びその制御方法

(57) 【要約】

【目的】 記憶装置システム内の制御装置と処理装置とを接続するデータ転送路の長さの許容範囲の拡大に伴って生じる、性能の低下、制御上の不具合等を低減した記憶装置システムを提供する。

【構成】 制御装置110と処理装置120の間で、データ転送を行う際に生ずる遅延時間を、データ転送路遅延時間測定記憶機能aにより、データ転送経路ごとに計測し、共有メモリ140上のデータ転送路遅延時間テーブル141に記憶する。一方、遅延時間対応制御機能aは、データ転送路遅延時間テーブル141を参照し、この記憶されているデータ転送遅延時間にしたがって、処理装置120から受け付けた入出力要求に対する制御を変更する。



【特許請求の範囲】

【請求項1】1台以上の記憶装置と制御装置を含み、1本以上のデータ転送路により1台以上の処理装置に接続された記憶装置システムであって、

前記データ転送路のそれぞれに対し、前記データ転送路をデータが伝搬される際に生ずる遅延時間に関する情報を計測して記憶する手段と、

記憶した前記遅延時間に関する情報にしたがって、前記処理装置から前記記憶装置に対して発行された入出力要求に対する制御を変更する手段を有することを特徴とする記憶装置システム。

【請求項2】データ転送に先立ち位置付け処理が必要な1台以上の記憶装置と、制御装置を含み、1本以上のデータ転送路により1台以上の処理装置に接続され、前記記憶装置が前記位置付け処理を完了した後、前記記憶装置に対し前記処理装置から受け付けた入出力要求に対し前記データ転送路を占有させる占有処理を実行する記憶装置システムであって、

前記データ転送路のそれぞれに対し、前記データ転送路をデータが伝搬される際に生ずる遅延時間に関する情報を計測して記憶する手段と、

記憶した前記データ転送路の前記遅延時間に関する情報にしたがって、前記占有処理の実行開始時間を変更する手段を有することを特徴とする記憶装置システム。

【請求項3】請求項2記載の記憶装置システムにおいて、前記実行開始時間を変更する手段は、前記入出力要求が使用することを許されているデータ転送路に関して、記憶した前記遅延時間に関する情報の最大値を基に、前記占有処理の実行開始時間を変更することを特徴とする記憶装置システム。

【請求項4】請求項2記載の記憶装置システムにおいて、前記実行開始時間を変更する手段は、前記入出力要求が使用することを許されているデータ転送路に関して、記憶した前記遅延時間に関する情報の平均値を基に、前記占有処理の実行開始時間を変更することを特徴とする記憶装置システム。

【請求項5】1台以上の記憶装置と制御装置を含み、1本以上のデータ転送路により複数の処理装置に接続された記憶装置システムであって、
前記データ転送路のそれぞれに対し、前記データ転送路をデータが伝搬される際に生ずる遅延時間に関する情報を計測して記憶する手段と、
前記記憶装置に対して、前記処理装置から受け付けた入出力要求の処理を完了させた時、前記記憶装置に対して受け付け待ち状態にしていた他の複数の入出力要求の待ちの解除処理の優先付けを、待ち状態にしていた前記入出力要求が使用することを許されているデータ転送路に関して記憶した前記遅延時間に関する情報に基づいて変更する手段を有することを特徴とする記憶装置システム。

【請求項6】請求項5記載の記憶装置システムにおい

て、前記記憶装置に対して、前記処理装置から受け付けた入出力要求の処理を完了させた時、前記記憶装置に対して受け付け待ち状態にしていた他の複数の入出力要求の待ちの解除処理の優先付けを、待ち状態にしていた前記入出力要求が使用することを許されているデータ転送路に関して、記憶した前記遅延時間に関する情報の平均値、あるいは、最小値にしたがって、変更する手段を有することを特徴とする記憶装置システム。

【請求項7】1台以上の記憶装置と制御装置を含み、1本以上のデータ転送路により1台以上の処理装置に接続され、前記処理装置のそれぞれから前記記憶装置のそれぞれに入出力要求を発行する際、使用が許可されている前記データ転送路の集合が定められた記憶装置システムであって、

前記データ転送路のそれぞれに対し、前記データ転送路をデータが伝搬される際に生ずる遅延時間に関する情報の計測と記憶を行い、前記処理装置のそれぞれから前記記憶装置のそれぞれに入出力要求を発行する際、前記使用が許可されているデータ転送路の集合ごとに、記憶した前記遅延時間に関する情報から統計情報を作成して記憶する手段と、

前記統計情報にしたがって、前記処理装置から前記記憶装置に発行された入出力要求に対する制御を変更する手段とを有することを特徴とする記憶装置システム。

【請求項8】データ転送に先立ち位置付け処理が必要な1台以上の記憶装置と制御装置を含み、1本以上のデータ転送路により1台以上の処理装置に接続され、前記記憶装置が前記位置付け処理を完了した後、前記記憶装置に対し前記処理装置から受け付けた入出力要求に対し前記データ転送路を占有させる占有処理を実行し、前記処理装置のそれぞれから前記記憶装置のそれぞれに入出力要求を発行する際、使用が許可されている前記データ転送路の集合が定められた記憶装置システムであって、
前記データ転送路のそれぞれに対し、前記データ転送路をデータが伝搬される際に生ずる遅延時間に関する情報の計測と記憶を行い、前記処理装置のそれぞれから前記記憶装置のそれぞれに入出力要求を発行する際に使用が許可されている前記データ転送路の前記集合ごとに、記憶した前記遅延時間に関する情報から統計情報を作成して記憶する手段と、
前記統計情報にしたがって、前記占有処理の実行開始時間を変更する手段とを有することを特徴とする記憶装置システム。

【請求項9】請求項8記載の記憶装置システムにおいて、前記記憶する手段は、前記統計情報として、前記処理装置のそれぞれから前記記憶装置のそれぞれに入出力要求を発行する際に使用が許可されている前記データ転送路の前記集合の前記遅延時間の最大値を記憶することを特徴とする記憶装置システム。

【請求項10】請求項9記載の前記記憶装置システムに

において、前記実行開始時間を変更する手段は、記憶した前記遅延時間の最大値にしたがって、前記占有処理の実行開始時間を変更することを特徴とする記憶装置システム。

【請求項 1 1】 1 台以上の記憶装置と制御装置を含み、1 本以上のデータ転送路により複数の処理装置に接続され、前記処理装置のそれぞれから前記記憶装置それぞれに入出力要求を発行する際、使用が許可されている前記データ転送路の集合が定められた記憶装置システムであって、

前記データ転送路のそれぞれに対し、前記データ転送路をデータが伝搬される際に生ずる遅延時間に関する情報の計測と記憶を行い、前記処理装置のそれぞれから前記記憶装置のそれぞれに入出力要求を発行する際、使用が許可されている前記データ転送路の前記集合ごとに、記憶した前記遅延時間に関する情報から統計情報を作成して記憶する手段と、

前記記憶装置に対して、前記処理装置から受け付けた入出力要求の処理を完了させた時、記憶した前記統計情報にしたがって前記記憶装置に対して受付け、待ち状態にしていた他の複数の入出力要求の待ちの解除処理の優先付けを変更する手段を有することを特徴とする記憶装置システム。

【請求項 1 2】 請求項 1 1 記載の記憶装置システムにおいて、前記記憶する手段は前記統計情報として、前記処理装置のそれぞれから前記記憶装置のそれぞれに入出力要求を発行する際、使用が許可されている前記データ転送路の前記集合の前記遅延時間の平均値、あるいは、最小値を記憶し、前記変更する手段は、前記記憶装置に対して、前記処理装置から受け付けた入出力要求の処理を完了させた時、記憶した前記遅延時間の前記平均値、あるいは、前記最小値に基づいて、前記記憶装置に対して受付け、待ち状態にしていた他の複数の入出力要求の待ちの解除処理の優先付けを変更することを特徴とする記憶装置システム。

【請求項 1 3】 1 台以上の記憶装置と制御装置を含み、1 本以上のデータ転送路により 1 台以上の処理装置に接続された記憶装置システムであって、前記データ転送路のそれぞれに対し、前記データ転送路をデータが伝搬される際に生ずる遅延時間に関する情報の計測と記憶を行い、前記データ転送路が接続された前記処理装置ごとに、記憶した前記遅延時間に関する情報から統計情報を作成して記憶する手段と、記憶した前記遅延時間に関する情報の前記統計情報にしたがって、前記処理装置から前記記憶装置に発行された入出力要求に対する制御を変更する手段を有することを特徴とする記憶装置システム。

【請求項 1 4】 データ転送に先立ち位置付け処理が必要な 1 台以上の記憶装置と制御装置を含み、1 本以上のデータ転送路により 1 台以上の処理装置に接続され、前記

記憶装置が前記位置付け処理を完了した後、前記記憶装置に対し前記処理装置から受け付けた入出力要求に対し前記データ転送路を占有させる占有処理を実行する記憶装置システムであって、

前記データ転送路のそれぞれに対し、前記データ転送路をデータが伝搬される際に生ずる遅延時間に関する情報の計測と記憶を行い、前記データ転送路が接続された前記処理装置ごとに、記憶した前記遅延時間に関する情報から統計情報を作成して記憶する手段と、

記憶した前記データ転送路の前記遅延時間に関する情報の前記統計情報にしたがって、前記占有処理の実行開始時間を変更する手段を有することを特徴とする記憶装置システム。

【請求項 1 5】 請求項 1 4 記載の記憶装置システムにおいて、前記記憶する手段は、前記統計情報として、前記データ転送路が接続された前記処理装置ごとに前記遅延時間の最大値を記憶することを特徴とする記憶装置システム。

【請求項 1 6】 請求項 1 5 記載の記憶装置システムにおいて、前記変更する手段は、記憶した前記遅延時間の最大値にしたがって、前記占有処理の実行開始時間を変更することを特徴とする記憶装置システム。

【請求項 1 7】 1 台以上の記憶装置と制御装置を含み、1 本以上のデータ転送路により複数の処理装置に接続した記憶装置システムであって、

前記データ転送路のそれぞれに対し、前記データ転送路をデータが伝搬される際に生ずる遅延時間に関する情報の計測と記憶を行い、前記データ転送路が接続された前記処理装置ごとに、記憶した前記遅延時間に関する情報から統計情報を作成して記憶する手段と、

前記記憶装置に対して、前記処理装置から受け付けた入出力要求の処理を完了させた時、記憶した前記遅延時間に関する情報の前記統計情報にしたがって、前記記憶装置に対して受付け、待ち状態にしていた他の複数の入出力要求の待ちの解除処理の優先付けを変更する手段を有することを特徴とする記憶装置システム。

【請求項 1 8】 請求項 1 1 記載の記憶装置システムにおいて、前記記憶する手段は、前記統計情報として、前記データ転送路が接続された前記処理装置ごとに、前記遅延時間の平均値、あるいは、最小値を記憶し、前記変更する手段は、前記記憶装置に対して、前記処理装置から受け付けた入出力要求の処理を完了させた時、記憶した前記遅延時間の前記平均値、あるいは、前記最小値にしたがって、前記記憶装置に対して受付、待ち状態にしていた他の複数の入出力要求の待ちの解除処理の優先付けを変更する手段を有することを特徴とする記憶装置システム。

【請求項 1 9】 1 台以上の記憶装置と制御装置を含み、1 本以上のデータ転送路により 1 台以上の処理装置に接続された記憶装置システムの制御方法であって、

前記データ転送路のそれぞれに対し、前記データ転送路をデータが伝搬される際に生ずる遅延時間に関する情報を計測して記憶するステップと、

記憶した前記遅延時間に関する情報にしたがって、前記処理装置から前記記憶装置に対して発行された入出力要求に対する制御を変更するステップを有することを特徴とする記憶装置システムの制御方法。

【請求項20】データ転送に先立ち位置付け処理が必要な1台以上の記憶装置と制御装置を含み、1本以上のデータ転送路により1台以上の処理装置に接続され、前記記憶装置が前記位置付け処理を完了した後、前記記憶装置に対し前記処理装置から受け付けた入出力要求に対し前記データ転送路を占有させる占有処理を実行する記憶装置システムの制御方法であって、前記データ転送路のそれぞれに対し、前記データ転送路をデータが伝搬される際に生ずる遅延時間に関する情報を計測して記憶するステップと、記憶した前記データ転送路の前記遅延時間に関する情報にしたがって、前記占有処理の実行開始時間を変更するステップを有することを特徴とする記憶装置システムの制御方法。

【請求項21】1台以上の記憶装置と制御装置を含み、1本以上のデータ転送路により複数の処理装置に接続された記憶装置システムの制御方法であって、前記データ転送路のそれぞれに対し、前記データ転送路をデータが伝搬される際に生ずる遅延時間に関する情報を計測して記憶するステップと、前記記憶装置に対して、前記処理装置から受け付けた入出力要求の処理を完了させた時、待ち状態にしていた前記入出力要求が使用することを許されているデータ転送路に関して、記憶した前記遅延時間に関する情報にしたがって、前記記憶装置に対して受付け、待ち状態にしていた他の複数の入出力要求の待ちの解除処理の優先付けを変更するステップを有することを特徴とする記憶装置システムの制御方法。

【請求項22】1台以上の記憶装置と制御装置を含み、1本以上のデータ転送路により1台以上の処理装置に接続され、前記処理装置のそれぞれから前記記憶装置のそれぞれに入出力要求を発行する際、使用が許可されている前記データ転送路の集合が定められた記憶装置システムの制御方法であって、前記データ転送路のそれぞれに対し、前記データ転送路をデータが伝搬される際に生ずる遅延時間に関する情報の計測と記憶を行い、前記処理装置のそれぞれから前記記憶装置のそれぞれに入出力要求を発行する際、使用が許可されている前記データ転送路の前記集合ごとに、記憶した前記遅延時間に関する情報から統計情報を作成して記憶するステップと、記憶した前記遅延時間に関する情報の前記統計情報にしたがって、前記処理装置から前記記憶装置に発行された

入出力要求に対する制御を変更するステップを有することを特徴とする記憶装置システムの制御方法。

【請求項23】データ転送に先立ち位置付け処理が必要な1台以上の記憶装置と制御装置を含み、1本以上のデータ転送路により1台以上の処理装置に接続され、前記記憶装置が前記位置付け処理を完了した後、前記記憶装置に対し前記処理装置から受け付けた入出力要求に対し前記データ転送路を占有させる占有処理を実行し、前記処理装置のそれぞれから前記記憶装置のそれぞれに入出力要求を発行する際、使用が許可されている前記データ転送路の集合が定められた記憶装置システムの制御方法であって、

前記データ転送路のそれぞれに対し、前記データ転送路をデータが伝搬される際に生ずる遅延時間に関する情報の計測と記憶を行い、前記処理装置のそれぞれから前記記憶装置のそれぞれに入出力要求を発行する際、使用が許可されている前記データ転送路の前記集合ごとに、記憶した前記遅延時間に関する情報から統計情報を作成して記憶するステップと、

記憶した前記データ転送路の前記遅延時間に関する情報の前記統計情報にしたがって、前記占有処理の実行開始時間を変更するステップを有することを特徴とする記憶装置システムの制御方法。

【請求項24】1台以上の記憶装置と制御装置を含み、1本以上のデータ転送路により複数の処理装置に接続され、前記処理装置のそれぞれから前記記憶装置のそれぞれに入出力要求を発行する際、使用が許可されている前記データ転送路の集合が定められた記憶装置システムの制御方法であって、

前記データ転送路のそれぞれに対し、前記データ転送路をデータが伝搬される際に生ずる遅延時間に関する情報の計測と記憶を行い、前記処理装置のそれぞれから前記記憶装置のそれぞれに入出力要求を発行する際、使用が許可されている前記データ転送路の前記集合ごとに、記憶した前記遅延時間に関する情報から統計情報を作成して記憶するステップと、

前記記憶装置に対して、前記処理装置から受け付けた入出力要求の処理を完了させた時、記憶した前記遅延時間に関する情報の前記統計情報にしたがって、前記記憶装置に対して受付け、待ち状態にしていた他の複数の入出力要求の待ちの解除処理の優先付けを変更するステップを有することを特徴とする記憶装置システムの制御方法。

【請求項25】1台以上の記憶装置と制御装置を含み、1本以上のデータ転送路により1台以上の処理装置に接続された記憶装置システムの制御方法であって、前記データ転送路のそれぞれに対し、前記データ転送路をデータが伝搬される際に生ずる遅延時間に関する情報の計測と記憶を行い、前記データ転送路が接続された前記処理装置ごとに、記憶した前記遅延時間に関する情報

から統計情報を作成して記憶するステップと、記憶した前記遅延時間に関する情報の前記統計情報にしたがって、前記処理装置から前記記憶装置に発行された入出力要求に対する制御を変更するステップを有することを特徴とする記憶装置システムの制御方法。

【請求項26】データ転送に先立ち位置付け処理が必要な1台以上の記憶装置と制御装置を含み、1本以上のデータ転送路により1台以上の処理装置に接続され、前記記憶装置が前記位置付け処理を完了した後、前記記憶装置に対し前記処理装置から受け付けた入出力要求に対し前記データ転送路を占有させる占有処理を実行する記憶装置システムの制御方法であって、前記データ転送路のそれぞれに対し、前記データ転送路をデータが伝搬される際に生ずる遅延時間に関する情報の計測と記憶を行い、前記データ転送路が接続された前記処理装置ごとに、記憶した前記遅延時間に関する情報から統計情報を作成して記憶するステップと、記憶した前記データ転送路の前記遅延時間に関する情報の前記統計情報にしたがって、前記占有処理の実行開始時間を変更するステップを有することを特徴とする記憶装置システムの制御方法。

【請求項27】1台以上の記憶装置と制御装置を含み、1本以上のデータ転送路により複数の処理装置に接続した記憶装置システムの制御方法であって、前記データ転送路のそれぞれに対し、前記データ転送路をデータが伝搬される際に生ずる遅延時間に関する情報の計測と記憶を行い、前記データ転送路が接続された前記処理装置ごとに、記憶した前記遅延時間に関する情報から統計情報を作成して記憶するステップと、前記記憶装置に対して、前記処理装置から受け付けた入出力要求の処理を完了させた時、記憶した前記遅延時間に関する情報の前記統計情報にしたがって、前記記憶装置に対して受付け、待ち状態にしていた他の複数の入出力要求の待ちの解除処理の優先付けを変更するステップを有することを特徴とする記憶装置システムの制御方法。

【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明は、計算機システム等に用いられる記憶装置システムに係り、特に、処理装置と接続可能距離が大きいケーブルを介して接続された記憶装置システムに関する。

【0002】

【従来の技術】計算機システム等における処理装置と記憶装置間を接続する転送路のアーキテクチャに関する技術として、例えば、日経コンピュータ 1990年9月24日号の第57～58頁に記載されるような技術が知られている。本文献には、光ケーブルを用いて処理装置と記憶装置との間の接続を行なう記憶装置システムに関する技術が開示されている。このように、光ケーブルを

用いて処理装置と記憶装置とを接続することにより、接続可能距離をkmのオーダーへと飛躍的に拡大することが可能となる。

【0003】一方、日立マニュアル HITAC H-6581-C3形ディスク制御装置H-6587形ディスク駆動装置 H-6586形ディスク駆動装置 H-6585形ディスク駆動装置 H-8589形ディスク駆動装置 8080-2-130-10の第138頁には、計算機システムに用いられる記憶装置の代表的な1つであるディスク装置のデータ転送方法が記載されている。

【0004】通常、ディスク装置は、データ転送に先立ち、シーク処理、回転待ち処理等の位置付け処理の実行が必要となる。この位置付け処理の実行中、制御装置と処理装置の間のデータ転送路、制御装置とディスク装置の間のデータ転送路は開放される。この後、ディスク装置のヘッドが、リード/ライトしたいデータの記録位置に近づいた時、それまで開放していたデータ転送路を再び占有し（この処理を以下、再接続処理と呼ぶ）、データのリード/ライト処理の実行に入る。

【0005】

【発明が解決しようとする課題】上記従来技術では、光ケーブルを用いることにより、処理装置と記憶装置の間の接続距離を拡大させている。処理装置と記憶装置との間の接続距離を拡大させることにより、従来は大きな問題とならなかった、データがケーブルを伝搬するのに要する時間、即ち伝搬遅延時間が、記憶装置とのデータ転送時間に比較して無視できない値となる。しかしながら、上述した従来技術ではこのような伝搬遅延時間については、何等考慮されていなかった。

【0006】例えば、従来のように、処理装置とディスク装置とを接続するケーブルの長さが一様に短かった状況下では、ディスク装置の位置付け処理後に行なわれる再接続処理の実行開始のタイミングは、ディスク装置のヘッドが、リード/ライトするデータの記録位置から一定距離分だけ、前に位置付いた時であった。しかし、処理装置とディスク装置を接続するケーブルの長さが長くなると、処理装置まで再接続処理の実行開始を通知するのに要する時間が長くなり、再接続処理が完了したときにはディスク装置のヘッドがリード/ライトの対象となるデータの記録位置を通過してしまうという事態が生じる。このため、再度データの記録位置にヘッドが位置付くのを待たねばならなくなり、性能が大幅に低下してしまうという問題が発生する。

【0007】さらに、複数の処理装置から同一の記憶装置への入出力要求が競合した時に生ずる不具合の例がある。例えば、記憶装置が、ある処理装置から受けた入出力要求に対する処理を実行中に、他の2つ以上の処理装置から別の入出力要求を受付け、それらを待たせた状態で実行中の入出力要求の処理が完了したことを考えてみ

る。このような場合、通常、記憶装置は、各処理装置に対して次の入出力要求が実行可能になったことを各処理装置に対して通知し、最も早く応答のあった処理装置の入出力要求の処理を実行している。従来のように、各処理装置と記憶装置とを接続するケーブルの長さが一樣に短かった状況下では、このように、処理装置からの応答が最も速かった入出力要求を、次に実行しても特に問題となることはなかった。しかし、ケーブル長の範囲が拡大し、各処理装置と記憶装置との間のケーブル長が異なると、記憶装置から送られる次の入出力要求の実行可能を知らせる通知が各処理装置に到達するまでの時間に差が生じる。このため、従来のような方式では、記憶装置との間の接続距離が長い処理装置の入出力要求は、常に待たされ、サービスを受けることができない、といった状態に陥る恐れがある。

【0008】従って、本発明の目的は、処理装置と記憶装置との間の接続可能距離の延長に伴うデータの伝搬遅延時間の増大によって発生する、記憶装置の制御上の上述したような不具合を解決した記憶装置システムを提供することにある。

【0009】

【課題を解決するための手段】本発明による記憶装置システムは、上記課題を達成するために、記憶装置と処理装置との間のデータ転送を制御するための制御装置に、各データ転送路ごとにその長さ、あるいは、データがその転送路を伝搬するのに要する遅延時間を計測し、この計測値を制御装置内の内部メモリに記憶するデータ転送路遅延時間計測記憶機構と、記憶した各データ転送路の遅延時間の計測値を参照し、この計測値にしたがって、処理装置から受け付けた入出力要求に対する制御を変更する遅延時間制御機構を有する。

【0010】本発明の記憶装置システムは、より好ましい態様において、遅延時間制御機構に、再接続処理の実行開始契機を、この計測値にしたがって変更する再接続開始契機変更機構を有する。また、記憶装置に対する複数の入出力要求を解除する際、長いデータ転送路を通じて受け付けた入出力要求に対するサービスの保証を行なうため、データ転送路の長さ等の計測値を参照し、長いデータ転送路を通じて受け付けた入出力要求の待ち状態を優先的に解除する待ち状態優先解除機構を有する。

【0011】

【作用】本発明では、上述したデータ転送路遅延時間計測記憶機構によって、処理装置との間に設けた各ケーブル、すなわち、各データ転送路ごとの長さ、あるいは、その転送路をデータが伝搬される際の遅延時間を計測し、その計測値を制御装置内の内部メモリに記憶する。さらに、上述した遅延時間制御機構により、記憶した各データ転送路の計測値を参照し、この計測値にしたがって、処理装置から受け付けた入出力要求に対する制御を決定する。これによって、データがケーブルを伝搬され

るのに要するオーバーヘッド時間が増大したことによって生ずる、記憶装置の制御上の不具合を解決することが可能となる。

【0012】より具体的には、例えば、再接続開始契機変更機構により、データ転送路遅延時間計測記憶機構における計測値にしたがって、再接続処理の実行開始契機を変更する。これによって、再接続処理が完了した時には、ディスク装置のヘッドが、リード/ライトしたいデータを通過してしまい、大幅に性能が落ちるという問題を回避できる。また、待ち状態優先解除機構により、データ転送路遅延時間計測記憶機構における計測値にしたがって、長いデータ転送路を通じて受け付けた入出力要求この入出力要求の待ち状態を優先的に解除する。これによって、長いデータ転送路を通じて受け付けた入出力要求に対するサービスの保証を行なうことも可能となる。

【0013】

【実施例】以下、本発明について、図面を用い詳細に説明する。

【0014】図1は本発明を適用した計算機システムの第1の実施例の概要を示す構成図である。

【0015】図1において、100は記憶装置システムであり、制御装置110と複数の記憶装置150を含んで構成される。制御装置110は、1つ以上（本実施例では4つ）の転送制御プロセッサ130、共有メモリ140、処理装置側インターフェース160、及び記憶装置側インターフェース170を有している。また、処理装置120は、制御装置110とのインターフェイスとして1つ以上（本実施例では2つ）のチャンネル121を有している。チャンネル121は、光ファイバ等のケーブルによって制御装置110の処理装置側インターフェース160に接続される。なお、処理装置120がチャンネル121を持たず、記憶装置システム100と直接ケーブルで接続されても構わない。同様に、制御装置110においても、処理装置側インターフェース160を設けることなく、転送制御プロセッサ130に直接ケーブル190を接続するような構成を採ることもできる。また、各処理装置120と制御装置110を接続するケーブル190は、それぞれ長さが異なってもよい。さらには、同一の処理装置120と制御装置110の間に接続されたケーブル190の長さが互いに異なることも許容される。

【0016】各転送制御プロセッサ130は、本発明に関係して有する機能として、データ転送路遅延時間測定記憶機能a131と遅延時間対応制御機能a132を備えており、記憶装置150と処理装置120との間のデータ転送を制御する。本実施例では、データ転送路遅延時間測定記憶機能a131と遅延時間対応制御機能a132は、転送制御プロセッサ130のプログラム制御による機能として実現されるが、専用のハードウェア等

同等の機能を実現し、制御装置110に組み込んでよい。

【0017】共有メモリ140は、各転送制御プロセッサ130が参照する共有情報を格納する。共有メモリ140には、データ転送路遅延時間テーブル141等の情報が格納されている。

【0018】制御装置110の各転送制御プロセッサ130は、記憶装置側インターフェイス170を経由して記憶装置150と接続される。もちろん、記憶装置インターフェイス170を経由せず、直接転送制御プロセッサ130と記憶装置130を接続することも可能である。

【0019】ここで、ある処理装置120と各記憶装置150の間のデータ転送処理を行う場合、その処理装置120と制御装置110の間に設置されたすべてのケーブル190の使用が許されてもよく、または、一部のケーブル190の使用しか許さなくてもよい。また、使用可能なケーブル190は、記憶装置150ごとに異なるようにしてもよい。

【0020】図2は、データ転送の遅延を計測する単位となるデータ転送路として、2つの例を示している。1つは、図2に示すデータ転送路a201のように、処理装置120と間に設置された各ケーブル190に相当する部分である。もう1つは、図2に示すデータ転送路b202のように、処理装置110から制御装置110内の転送制御プロセッサ130までの経路に相当する部分である。データ転送の遅延を計測する単位としては、これらいずれの単位を用いてもよい。

【0021】図3は、共有メモリ140内に設けられるデータ転送路遅延時間テーブル141の構造を表している。データ転送路遅延時間テーブル141には、各データ転送路200に対応してデータ転送路遅延時間300が格納される。具体的には、対応するデータ転送路200をデータが伝搬された時に生ずる遅延時間が設定される。本実施例では、データ転送遅延時間300としてデータ転送路200をデータが伝搬される際の遅延時間そのものを用いるが、他に、データ転送路200の長さ等、遅延時間を計算、あるいは、予測可能なものであれば何を用いてもよい。

【0022】データ転送路遅延時間テーブル140のエントリの数は、図2に示すデータ転送路a201を計測単位とする場合、接続しているケーブル190の数であり、データ転送路b202を計測単位とする場合、最大で、接続しているケーブル190の数に転送制御プロセッサ130の数を乗じた値となる。

【0023】次に、転送制御プロセッサ130によって実現されるデータ転送路遅延時間測定記憶機能a131の動作について説明する。処理装置110が制御装置120に入出力要求を発行する際、その入出力要求の内容を規定する必要がある。この内容は、通常、コマンドと

呼ばれる指示単位の組合せで規定される。1つのコマンドを処理装置110から制御装置120に送るには、複数の制御情報を交換することがプロトコルとして定められている。本実施例では、データ転送路200の遅延時間の測定には、特定のコマンドを用いている。もちろん、遅延時間を測定するコマンドを特定せず、任意の複数のコマンドや、全てのコマンドについて測定を行なってもよい。

【0024】制御装置110は、特定したコマンドで定められたプロトコルに従って、処理装置120へ、処理装置120からの応答を返す制御情報を送信し、処理装置120からその応答として制御情報を受信するまでの時間を遅延時間として測定する。測定の方法としては、何回かの制御情報の送受信時間を計測、制御情報の送信時間、受信時間の一方だけ、あるいは、双方を別々に計測などの方法を用いることができる。

【0025】図4は、データ転送路遅延時間測定記憶機能a131によって実施される処理のフローを示すフローチャートである。以下、図4を用いこの処理フローの説明を行う。データ転送路遅延時間測定記憶機能a131は、転送制御プロセッサ130により起動されると、ステップ400で遅延時間を測定すべきコマンドか否かを調べ、遅延時間を測定すべきコマンドでなければ何もしずリターンする。遅延時間を測定するコマンドならば、ステップ410で、制御プロセッサ130が処理装置110へ情報転送するまで待ち、ステップ420で、タイマをスタートさせる。その後、データ転送路遅延時間測定記憶機能a131は、ステップ430で、処理装置120からデータを受領するまで待ち、データを受領した後ステップ440で、タイマをストップさせる。最後に、ステップ450で、この時のタイマの値から、経過時間を求め、それをデータ転送路遅延時間テーブル141内の、処理中のデータ転送路200に対応するエントリにデータ転送路遅延時間300として格納する。

【0026】図5は、遅延時間対応制御機能a132により実施される処理のフローを示すフローチャートである。遅延時間対応制御機能a132では、ステップ500で、まず、データ転送路遅延時間テーブル141の、起動チャンネル121に対応するエントリから、データ転送路遅延時間300を取り出す。そして、ステップ510では、この取り出したデータ転送路遅延時間300の値にしたがって、処理中の入出力要求に対する制御を決定する。

【0027】図6は、遅延時間対応制御機能a132が有する具体的な機能と、その機能に關係して共有メモリ140に設けられる情報を表している。

【0028】遅延時間対応制御機能a132は、再接続開始時間調節機構a600と優先待ち状態解除機構a601を有している。共有メモリには、優先待ち状態解除機構a601に關係して、待ち状態入出力要求登録デー

ブル602が設けられる。なお、データ転送路遅延時間テーブル141は、再接続開始時間調節機構a600と優先待ち状態解除機構a601の双方から参照される。再接続開始時間調節機構a600が動作するのは、記憶装置150に対し、再接続処理を実行する必要がある場合に限られる。以下、再接続開始時間調節機構a600の説明に先立って、再接続処理自体についての説明を行なう。

【0029】図7は、本実施例において記憶装置150として用いるディスク装置の構成を示す構成図である。

【0030】記憶装置150は、1枚以上の円形の記憶媒体740、ヘッド750、記憶媒体インタフェース730、記憶媒体制御部710、及びセクタ検出部720から構成される。記憶媒体制御部710は、記憶媒体740と制御装置110の間のデータ転送を制御する。ヘッド750は、回転している記憶媒体740からデータを読み書きする機能を持ち、記憶媒体740単位に存在する。

【0031】円形の記憶媒体740は、一定の角度ごとにセクタ760という単位が設定される。各セクタ760を一意に認識するため、各セクタ760には、番号がつけられる。セクタ検出部720は、ヘッド750が現在読み書きできるセクタ760を認識する機能をもつ。さらに、セクタ検出部720は、制御装置110から指定されたセクタ760にヘッド750が到達した時、記憶媒体制御部710経由で制御装置110へ割込みを上げる機能をもつ。

【0032】記憶媒体740が1回転する間に、ヘッド750がアクセス可能な円状の記録単位をトラック741といい、記憶媒体740上に1つ以上存在する。あるトラック741が入出力対象になったときに、ヘッド750はそのトラック741を読み書きできる位置に移される。この処理をシーク処理と呼ぶ。

【0033】図8に、1本のトラック741の典型的な構成例を示す。

【0034】トラック741のある位置に、インデクス820とよばれるトラック741の先頭が定められている。トラック741も、記憶媒体740上に含まれるため、先に説明したセクタ760に分割されていることになる。

【0035】トラック741には、読出しまたは書込みの単位となるレコード810が複数存在する。通常、各レコード810の認識を可能にするため、レコード810には、レコード番号が付けられている。

【0036】レコード810を読み書きするためには、シーク処理が完了した後、ヘッド750がそのレコード810を読み書きできるようになるまで、記憶媒体740が回転してくるのを待たなければならない。この処理を、回転待ち処理と呼ぶ。

【0037】シーク処理と回転待ち処理を併せて位置付

け処理と呼ぶ。記憶装置150が、位置付け処理を実行している間は、転送制御プロセッサ130、ケーブル190等の記憶装置150と処理装置120の間で、データ転送を行なうための資源を、確保したままにしておく必要はない。したがって、一般には、記憶装置150が位置付け処理を実行している間は、これらの資源を開放し、記憶装置の位置付け処理が完了した後、これらの資源を再確保して、データ転送処理を開始する。この資源の再確保のための処理を再接続処理と呼ぶ。

【0038】次に、図9を用いて、レコード810の一般的なアクセス例について説明する。処理装置120は、記憶装置150内のあるレコードをアクセスするために、チャンネル121より入出力要求900を制御装置110に送出する。入出力要求900は、記憶装置150上の読み書きするレコード810を決定するための情報として、トラック741の番号（トラック番号）、セクタ760の番号（セクタ番号）、レコード810の番号（レコード番号）を含んでいる。一般には、制御装置110に送るセクタの番号としては、読み書きしたいレコードが格納されているセクタの直前のセクタ番号が用いられる。

【0039】制御装置110内の転送制御プロセッサ130は、トラック番号、セクタ番号、レコード番号等を受け取ると、記憶装置150を確保し、記憶装置150に対し位置付け処理実行要求910を発行する。この時、転送制御プロセッサ130は、位置付け処理を完了し再接続処理を開始するセクタ760のセクタ番号を記憶装置150に伝える。再接続処理を開始するセクタとして指定するセクタ番号は、処理装置120から受け取ったセクタ番号そのものではなく、処理装置120から受け取ったセクタ番号に基づいて、転送制御プロセッサ130が計算して求めたものである。この記憶装置150におくるセクタ番号を決定するのが、再接続開始時間調節機構a600である。再接続開始時間調節機構a600の具体的な処理については、後述する。

【0040】記憶装置150は位置付け要求910を受領すると、まず、シーク処理を実行する。次に、セクタ検出部720を通じ、ヘッド750が、指定された番号のセクタ760に到達するのを待つ。目的のセクタ760に達すると、記憶装置150は位置付け処理が完了したことを割込みにより制御装置110に通知する。この割込みを、セットセクタ割込み940と呼ぶ。

【0041】セットセクタ割込みが発生した時、空いている（動作中でない）転送制御プロセッサ130が存在すれば、その転送制御プロセッサ130によりこのセットセクタ割込み940は受領され、処理装置120に再接続要求950が発行される。処理装置120が、再接続可能であれば、再接続要求950は受領され、処理装置120から再接続を要求した転送制御プロセッサ130に再接続許可通知960が発行される。転送制御プロ

セッサ130は、再接続許可通知960を受け取ると、他の転送制御プロセッサ130との間で、記憶装置150の競合を避けるため、自分が記憶装置150を占有することを通知する。以上の処理が、再接合処理における具体的な処理の内容である。

【0042】この後、記憶装置150上の目的のレコード810に対してリード/ライト処理が開始されることになる。

【0043】このように、最初に、記憶装置150がセットセクタ割込み940を通知し、再接続処理が完了するまでには、かなりの処理が実行される。この処理に要する時間は、記憶媒体740の回転時間に対して無視できない値となる。このため、セットセクタ割込み940は、実際にリード/ライトしたいレコード810がヘッド750を通過するより、ある程度前に通知する必要がある。従って、制御装置110では、セットセクタ割込み940を受けるためのセクタ番号として、処理装置120から受け取ったセクタ番号をそのまま用いることはできない。

【0044】従来のように、制御装置110と処理装置120とを接続するケーブル190の長さが一様に短く、データを交換する際の遅延時間がほとんど発生しない場合には、再接続処理に要する時間の変動要因はほとんどなかった。したがって、制御装置110は、セットセクタ割込み940を受けるためのセクタ番号として、処理装置120から受け取ったセクタ番号より、一定値だけ前のセクタ760のセクタ番号を用いればよかった。

【0045】しかし、ケーブル190の長さの許容範囲が拡大され、その長さにしたがって、処理装置120と制御装置110の間でデータを交換する際の遅延時間が変動するような場合には、再接続処理に要する時間が大きく変動することになる。したがって、セットセクタ割込み940を受けるためのセクタ番号として、処理装置120から受け取ったセクタ番号より一定のセクタ数だけ前のセクタ760のセクタ番号を用いると、再接続処理が完了した時には、ヘッド750が読み書きしたいレコード810を通過してしまい、あるいは、再接続処理が早く完了しすぎて、ヘッド450が読み書きしたいレコード810に位置づくまでに時間がかかりすぎる等、性能低下、転送制御プロセッサ130等の資源の無駄な占有といった問題が発生する。

【0046】再接続開始時間調節機構a600は、このような問題を解決するために、各データ転送路200の遅延時間にしたがって、セットセクタ割込み940を受けるためのセクタ760の番号を計算を行なう。再接続開始時間調節機構a600は、転送制御プロセッサ130が、記憶装置150に位置付け処理実行要求910を発行するに際して実行される。転送制御プロセッサ130は、再接続開始時間調節機構a600によって計算さ

れたセクタ番号を、セットセクタ割込み940を受けるためのセクタ番号として記憶装置150に通知する。

【0047】再接続開始時間調節機構a600の処理フローを図10に示す。転送制御プロセッサ130は、記憶装置150に位置付け要求910を発行する際、再接続開始時間調節機構a600を起動する。

【0048】再接続開始時間調節機構a600は、ステップ1000で、データ転送路遅延時間テーブル141から、処理中の入出力要求が使用を許されているデータ転送路200に対応するデータ転送路遅延時間300を取り出す。この後、再接続開始時間調節機構a600は、ステップ1010で、これらの値の最大値、あるいは、平均値をベース遅延値とする。どのデータ転送路200でも、再接続処理が間にあるようにする場合にはベース遅延値として最大値を用い、平均的な遅延時間のデータ転送路200において、再接続処理が間にあればよい場合には平均値を用いる。これらは、システムの要求に応じていずれか一方、あるいは両者を混在させて用いることができる。

【0049】ステップ1020では、再接続開始時間調節機構a600は、ステップ602で得た値をベース遅延値として、以下に示す数1にしたがって、セットセクタ割込みを受けるためのセクタ番号（割込み発生セクタ番号）を計算する。

【0050】割込み発生セクタ番号＝処理装置120から受け取ったセクタ番号－（再接続処理時の制御情報送受信回数×ベース遅延値＋オーバーヘッド時間）／セクタ通過時間
(数1)

数1の値が負の値、すなわち、トラック741の先頭を過ぎて、トラック741の後の方のセクタ760を選択する必要がある場合には、数1で得られた値に、トラック741内のセクタ760の数を加えた値を用いる。

【0051】転送制御プロセッサ130は、再接続開始時間調節機構a600によって計算されたセクタ番号を、位置付け要求を発行する際、記憶装置150に送る。

【0052】次に、優先待ち状態解除機構a601の説明を行う。

【0053】図11に、待ち状態入出力要求登録テーブル602の構成を示す。待ち状態入出力要求登録テーブル602は、入出力待ち情報1100により構成される。入出力待ち情報1100は、受け付けた入出力要求がリード/ライトの対象とする記憶装置150、及びその入出力要求を発行した処理装置120のそれぞれに対応して存在する。したがって、その数は、記憶装置150と処理装置120の積に等しい。制御装置110は、処理装置120から入出力要求を受け付けた時、アクセスしようとする記憶装置150が、他の入出力要求に占有されている場合には、その入出力要求を待ち状態にする。この時、この入出力要求を発行した処理装置12

0、及びアクセスしようとする記憶装置150に対応する入出力要求待ち情報1100に、待ち状態に入ったことを登録する。制御装置110が、各処理装置120から記憶装置150に発行された入出力要求を管理するための処理については、既知の技術を用いることができるので、特に詳細には説明しない。

【0054】優先待ち状態解除機構a601は、記憶装置150に対する複数の入出力要求を解除する際、データ転送路200のデータ転送路遅延時間300を参照し、長いデータ転送路200を通じて受け付けた入出力要求に対するサービスの保証を行なうため、このような入出力要求の待ち状態を優先的に解除する。

【0055】図12は、優先待ち状態解除機構a601の処理フローを示すフローチャートである。転送制御制御プロセッサ130は、ある記憶装置150に対する入出力要求が完了して、それまで待ち状態にしていた別の入出力要求の実行を開始する際に、優先待ち状態解除機構a601の機能を実行する。

【0056】ステップ1200で、優先待ち状態解除機構a601は、待ち状態入出力要求登録テーブル602に登録されている入出力要求の完了した記憶装置150に対応する入出力待ち情報1100をすべてサーチし、当該記憶装置150に対し待ち状態になっていた入出力要求をすべて認識する。ステップ1210では、データ転送路遅延時間テーブル141を参照し、それぞれの入出力要求が使用を許されているすべてのデータ転送路200に対応するデータ転送路遅延時間を取り出す。

【0057】次に、ステップ1220では、それぞれの入出力要求ごとに、すべてのデータ転送路遅延時間の最小値、あるいは、平均値を計算し、各入出力要求のベース遅延値とする。最も応答の速いデータ転送路200からの応答時間を重視する場合には、各入出力要求のベース遅延値として最小値を用い、平均的なデータ転送路の応答値を重視する場合には、平均値を用いる。これらは、システムの要求に応じていずれか一方、あるいは両者を混在させて用いることができる。

【0058】優先待ち状態解除機構a601は、ステップ1230で、ステップ1220において計算した、各入出力要求のベース遅延値を、降順にソーティングする。

【0059】ステップ1240では、ベース遅延値が大きい順に、入出力要求が実行可能になったことを、対応する処理装置120に、使用を許されているデータ転送路200を通じて報告する。ただし、この報告は、各入出力要求に対し、それぞれ前の入出力要求とのベース遅延値の差だけ間隔を置いて行なわれる。このように、各入出力要求に対応する処理装置120への報告に時間間隔を設けることにより、各処理装置120から、返ってくる入出力要求の再実行要求を受け取る時間をほぼ同一時間にすることができる。

【0060】このようにして、各処理装置に対して入出力要求の待ち状態の解除を報告した後、制御装置110は、入出力要求の再実行要求を受け取り、その処理を実行する。これらの処理については、従来のものと変わらないため、説明は省略する。

【0061】図13は、本発明を適用した計算機システムの第2の実施例の概要を示す構成図である。

【0062】本実施例と第1の実施例との主な相違は、本実施例では共有メモリ140上に、バスグループ対応遅延時間統計テーブル1300を含んでいる点である。なお、図2に示したデータ転送路200の考え方は、本実施例でもそのまま適用することができる。

【0063】第1の実施例では、遅延時間対応制御機能a132が制御を行なう際、処理中の入出力要求が発行された処理装置120等から、関係するすべてのデータ転送路200を認識し、対応するすべてのデータ転送路遅延時間300を取り出して、制御を行なった。しかし、制御装置130の処理は、極めて厳しい時間制約があり、第1の実施例では、処理オーバーヘッドが高い。そこで、本実施例では、データ転送路遅延時間測定記憶機能b1310に、データ転送路遅延時間テーブル141以外に、バスグループ対応遅延時間統計テーブル1300の情報を設定する機能を持たせる。そして、遅延時間対応制御機能b1320が動作する際には、バスグループ対応遅延時間統計テーブル1300の情報を参照した後述する処理によりオーバーヘッドを削減することを可能としている。

【0064】まず、バスグループ1330について説明する。第1の実施例でも説明したように、ある処理装置120が各記憶装置150とデータ転送処理を行う場合、その処理装置120と制御装置110の間に設置されたすべてのケーブル190の使用が許されているとは限らない。バスグループ1330とは、ある処理装置120がある記憶装置150とデータ転送処理を行う場合、使用が許可されているケーブル190の集合である。したがって、バスグループ1330は、処理装置120と記憶装置150の組合せの数だけ、定義されることになる。例えば、図13では、処理装置a1390は、チャンネルa1360、チャンネルb1370、チャンネルc1380を含むが、記憶装置a1340にアクセスする際に使用できるのは、チャンネルa1360とチャンネルb1370であり、記憶装置b1350にアクセスする際に使用できるのは、チャンネルb1370とチャンネルc1380である。したがって、バスグループa1331が、処理装置a1390が記憶装置a1340にアクセスする際に使用するバスグループ1330となり、バスグループb1332が、処理装置a1390が記憶装置b1350にアクセスする際に使用するバスグループ1330となる。

【0065】次に、簡単にデータ転送路200とバスグ

ループ1330との関係について説明しておく。データ転送路200を図2に示したデータ転送路a201とした場合、ケーブル190とデータ転送路200の関係は、1対1となる。したがって、データ転送路200が、どのバスグループ1330に属するかは一意に定められることになる。一方、データ転送路200を図2に示したデータ転送路b202のようにした場合、データ転送路200は、ケーブル190と転送制御プロセッサ130の組合せとなる。したがって、データ転送路200が、どのケーブル190に対応するかも明らかであるため、この場合も、データ転送路200が、どのバスグループ1330に属するかは一意に定められることになる。

【0066】図14には、バスグループ対応遅延時間統計テーブル1300の構成を示す。バスグループ対応遅延時間統計テーブル1300には、バスグループ1330対応して、そのバスグループ1330に含まれる各データ転送路200ごとに計測したデータ転送路遅延時間300に関して、統計処理を行った情報、即ち、バスグループ遅延時間統計値1400が記憶される。バスグループ遅延時間統計値1400には、バスグループ内のデータ転送路遅延時間の最大値1410、最小値1420、平均値1430等の情報が含まれる。この統計処理は、データ転送路遅延時間測定記憶機能b1310が実行する。本実施例のデータ転送路遅延時間測定記憶機能b1310と第1の実施例におけるデータ転送路遅延時間測定記憶機能a131とは、この統計処理を含む点で相違している。

【0067】図15は、データ転送路遅延時間測定記憶機能b1310の処理フローを示すフローチャートである。図15に示すフローチャートは、第1の実施例における、図4に示したデータ転送路遅延時間測定記憶機能a131の機能と共通部分が多い。図15に示したステップのうち、図4と同一の番号を付した処理ステップでは、図4を用いて説明した処理と同様の処理が行なわれる。

【0068】図15における本実施例の処理と図4に示す第1の実施例の処理の相違は、処理フローの最後で、バスグループ遅延時間統計値1400の更新（ステップ1500）を行なうか否かにある。

【0069】ステップ400乃至450は、第1の実施例において説明したデータ転送路の遅延時間測定のための処理と同様である。本実施例においてデータ転送路遅延時間測定記憶機能b1310は、各データ転送路についてデータ転送路遅延時間を求めた後、ステップ1500で、計測したデータ転送路200の遅延時間の値に対して統計処理を施す。そして、統計処理の結果としてバスグループ対応遅延時間統計値テーブル1300内の対応するバスグループ1330についてのバスグループ遅延時間統計値1400を更新する。

【0070】図16は、遅延時間対応制御機能b1320の処理フローを示すフローチャートである。本実施例における遅延時間対応制御機能b1320の処理は、データ転送路遅延時間テーブル141ではなく、バスグループ対応遅延時間統計テーブル1300の情報を参照して処理を行なう点で、第1の実施例における遅延時間対応制御機能a132と相違している。

【0071】遅延時間対応制御機能b1320は、ステップ1600で、処理中の入出力要求によるアクセスの対象となる記憶装置150と、入出力要求を発行した処理装置120から対応するバスグループ1330とを認識する。この結果、このバスグループ1330に対応するバスグループ遅延時間統計値1400をバスグループ対応遅延時間統計テーブル1300より取り出す。ステップ1610では、この取り出したバスグループ遅延時間統計値1400にしたがって、処理中の入出力要求に対する制御を変更する。

【0072】図17に、遅延時間対応制御機能b1320により実行される具体的機能と、その機能に関係して共有メモリ140に設けられる情報を示す。

【0073】遅延時間対応制御機能b1320は、その具体的機能として、再接続開始時間調節機構b1700と優先待ち状態解除機構b1710を有する。共有メモリに設けられる待ち状態入出力要求登録テーブル602は、第1の実施例で説明したものと同様のものが使用できる。

【0074】図18は、再接続開始時間調節機構b1700の処理フローを示すフローチャートである。本実施例において、再接続開始時間調節機構b1700により実現される処理は、データ転送路遅延時間テーブル141ではなく、バスグループ対応遅延時間統計テーブル1300の情報を参照して実行される点で、第1の実施例の再接続開始時間調節機構a600の処理と相違している。以下、再接続開始時間調節機構b1700により実行される処理について説明を行う。

【0075】再接続開始時間調節機構b1700は、ステップ1800で、処理中の入出力要求がアクセスの対象とする記憶装置150、及び処理中の入出力要求が発行された処理装置120から対応するバスグループ1330を認識する。この結果、このバスグループ1330に対応するバスグループ遅延時間統計値1400をバスグループ対応遅延時間統計テーブル1300から取り出す。この場合、合取り出す値としては、最大値1410、平均値1430のどちらを取り出して、使用してもよい。

【0076】次に、ステップ1020で、図10のステップ1020で説明したと同様に、数1に従って再接続時間を計算する。転送制御プロセッサ130は、再接続開始時間調節機構b1700による処理結果を用い、記憶装置150へ位置付け要求を発行する。

【0077】図19には、優先待ち状態解除機構b1710の処理フローを示す。本実施例における優先待ち状態解除機構b1710の処理は、データ転送路遅延時間テーブル141ではなく、バスグループ対応遅延時間統計テーブル1300の情報を参照して処理を実行する点で、第1の実施例の優先待ち状態解除機構a601の処理と相違する。以下、この説明を行う。

【0078】図19に示した各処理ステップのうち、図12と同一の番号を付した処理ステップでは、図12を用いて説明した処理と同様の処理が行なわれる。本実施例では、優先待ち状態解除機構b1710は、図12のステップ1210、1220に代えて、ステップ1900を実行する。他のステップにおける処理は、第1の実施例と同様であるので、詳細な説明は省略する。

【0079】優先待ち状態解除機構b1710は、ステップ1900において、ステップ1200で取り出した各入出力要求ごとに、アクセスの対象となる記憶装置150、及びその入出力要求を発行した処理装置120から、対応するバスグループ1330を認識する。この結果、各入出力要求ごとに、認識したバスグループ1330に対応するバスグループ遅延時間統計値1400をバスグループ対応遅延時間統計テーブル1300から取り出す。この場合取り出す値としては、最小値1420、平均値1430のどちらを取り出して使用してもよい。

【0080】これ以降、図12で説明したと同様に、ベース遅延値をソートし、その大きい順に入出力要求の待ち状態を解除する。

【0081】図20は、本発明を適用した計算機システムの第3の実施例の概要を示す構成図である。

【0082】本実施例では、共有メモリ140上に、第2の実施例で用いたバスグループ対応遅延時間統計テーブル1300に代えて、処理装置対応遅延時間統計テーブル2000を含んでいる点で上述した各実施例と相違する。なお、図2に示したデータ転送路200の考え方は、本実施例でもそのまま適用することができる。

【0083】第2の実施例では、遅延時間対応制御機能b1320のオーバーヘッドを削減するため、バスグループ対応遅延時間統計テーブル1300を用いた処理を行っている。しかし、バスグループ1330の数は、処理装置120と記憶装置150の組合せの数に等しくなるため、処理装置120と記憶装置150の組が増加すると、その分バスグループ対応遅延時間統計テーブル1300の容量が増大するという問題がある。そこで、本実施例では、各データ転送路200ごとに計測したデータ転送路遅延時間300を、データ転送路200を含むケーブル190が接続されている処理装置120に対応して統計情報化し、必要とする情報量を削減する。

【0084】図21に、処理装置対応遅延時間統計テーブル2000の構成を示す。処理装置対応遅延時間統計テーブル2000には、処理装置130に対応して処理

装置遅延時間統計値2100が記憶される。処理装置遅延時間統計値2100は、具体的には、その処理装置130と制御装置110との間の各データ転送路200ごとに計測したデータ転送路遅延時間300に関して、統計処理を行って得られた情報である。より具体的には、各処理装置との間の各データ転送路について得られたデータ転送路遅延時間300の最大値2110、最小値2120、平均値2130等の情報を含む。この統計処理は、データ転送路遅延時間測定記憶機能c2010により実行される。

【0085】図22は、データ転送路遅延時間測定記憶機能c2010の処理フローを示すフローチャートである。図22に示すフローチャートは、第2の実施例におけるデータ転送路遅延時間測定記憶機能b1310の処理フローと同様に、第1の実施例の説明に用いた図4に示すフローチャートと共通する部分が多い。図22に示したステップのうち、図4に示した処理ステップと同一の番号を付した処理ステップでは、図4を用いて説明したものと同様の処理が行なわれる。

【0086】図22における本実施例の処理と図4に示す第1の実施例の処理の相違は、処理フローの最後で、処理装置遅延時間統計値2100の更新（ステップ2200）を行なうか否かにある。

【0087】ステップ400乃至450は、第1の実施例において説明したデータ転送路の遅延時間測定のための処理と同様である。本実施例においてデータ転送路遅延時間測定記憶機能c2020は、各データ転送路についてデータ転送路遅延時間を求めた後、ステップ2200で、計測したデータ転送路200の遅延時間の値に対し、各処理装置に対応して統計処理を施す。そして、統計処理の結果として各処理装置に対応する処理装置対応遅延時間統計テーブル2000内の処理装置遅延時間統計値2100を更新する。

【0088】以下に説明する各処理は、処理装置対応遅延時間統計テーブル2000の情報を参照して行なわれる点で、上述した第1、及び第2の実施例と異なっている。

【0089】図23は、遅延時間対応制御機能c2020の処理フローを示すフローチャートである。遅延時間対応制御機能c2020は、ステップ2300で、処理中の入出力要求を発行した処理装置120に対応する処理装置遅延時間統計値2100を処理装置対応遅延時間統計テーブル2000から取り出す。ステップ2310では、この取り出した処理装置遅延時間統計値2100に応じて、処理中の入出力要求に対する制御を行なう。

【0090】図24には、遅延時間対応制御機能c2020が有する具体的機能と、その機能に関係して共有メモリ140に設けられる情報を示している。

【0091】遅延時間対応制御機能c2020は、具体的な機能として、再接続開始時間調節機構c2400と優

先待ち状態解除機構 c 2 4 1 0 を有する。共有メモリに設けられる待ち状態入出力要求登録テーブル 6 0 2 は、第 1 の実施例で説明したものと同様のものを用いることができる。

【0092】図 2 5 は、再接続開始時間調節機構 c 2 4 0 0 の処理フローを示すフローチャートである。再接続開始時間調節機構 c 2 4 0 0 は、ステップ 2 5 0 0 で、処理装置対応遅延時間統計テーブル 2 0 0 0 から処理中の入出力要求を発行した処理装置 1 2 0 に対応する処理装置遅延時間統計値 2 1 0 0 をとり出す。この場合取り出す値としては、最大値 2 1 1 0、平均値 2 1 3 0 のどちらを取り出して、使用してもよい。

【0093】次に、ステップ 1 0 2 0 で、図 1 0 のステップ 1 0 2 0 で説明したと同様に、数 1 に従って再接続時間を計算する。転送制御プロセッサ 1 3 0 は、再接続開始時間調節機構 c 2 4 0 0 による処理結果を用い、記憶装置 1 5 0 へ位置付け要求を発行する。

【0094】図 2 6 は、優先待ち状態解除機構 c 2 4 1 0 の処理フローを示すフローチャートである。図 2 6 に示した各処理ステップのうち、図 1 2 と同一の番号を付した処理ステップでは、図 1 2 を用いて説明した処理と同様の処理が行なわれる。本実施例では、優先待ち状態解除機構 c 2 4 1 0 は、図 1 2 のステップ 1 2 1 0、1 2 2 0 に代えて、ステップ 2 6 0 0 を実行する。他のステップにおける処理は、第 1 の実施例と同様であるので、詳細な説明は省略する。

【0095】優先待ち状態解除機構 c 2 4 1 0 は、ステップ 2 6 0 0 において、ステップ 1 2 0 0 で取り出した各入出力要求ごとに、その入出力要求を発行した処理装置 1 2 0 に対応する処理装置遅延時間統計値 2 1 0 0 を処理装置対応遅延時間統計テーブル 2 0 0 0 から取り出す。この場合取り出す値としては、最小値 2 1 2 0、平均値 2 1 3 0 のどちらを取り出して使用してもよい。

【0096】これ以降、第 1 の実施例において、図 1 2 で説明したと同様に、ベース遅延値をソートし、その大きい順に入出力要求の待ち状態を解除する。

【0097】

【発明の効果】本発明によれば、制御装置と処理装置の間を接続するデータ転送路の長さの許容範囲が拡大し、データ転送路をデータが伝搬される際に発生する遅延時間が記憶装置の制御に無視出来ない状態となった場合でも、遅延時間に応じた処理を行ない、データ転送路の無駄な占有の防止、あるいは、入出力要求の待ち状態を解除する際の各入出力要求の間のサービスバランスを図ることができる。

【図面の簡単な説明】

【図 1】本発明の第 1 の実施例の計算機システムの構成図。

【図 2】データ転送路の定義を説明するための概念図。

【図 3】データ転送路遅延時間テーブル 1 4 1 の構成

図。

【図 4】データ転送路遅延時間測定記憶機能 a の処理フローを示すフローチャート。

【図 5】遅延時間対応制御機能 a の処理フローを示すフローチャート。

【図 6】遅延時間対応制御機能 a 及び共有メモリに設けられる情報を示す概略構成図。

【図 7】記憶装置の構成を示す概略構成図。

【図 8】記憶媒体上のトラックの構成図。

【図 9】レコードの一般的なアクセス例を示すタイミングチャート。

【図 10】再接続開始時間調節機構 a の処理フローを示すフローチャート。

【図 11】待ち状態入出力要求登録テーブルの構成図。

【図 12】優先待ち状態解除機構 a の処理フローを示すフローチャート。

【図 13】本発明の第 2 の実施例の計算機システムの構成図。

【図 14】バスグループ対応遅延時間統計テーブルの構成図。

【図 15】データ転送路遅延時間測定記憶機能 b の処理フローを示すフローチャート。

【図 16】遅延時間対応制御機能 b の処理フローを示すフローチャート。

【図 17】遅延時間対応制御機能 b 及び共有メモリに設けられる情報を示す概略構成図。

【図 18】再接続開始時間調節機構 b の処理フローを示すフローチャート。

【図 19】優先待ち状態解除機構 b の処理フローを示すフローチャート。

【図 20】本発明の第 3 の実施例の計算機システムの構成図。

【図 21】処理装置対応遅延時間統計テーブルの構成図。

【図 21】データ転送路遅延時間測定記憶機能 c の処理フローを示すフローチャート。

【図 23】遅延時間対応制御機能 c の処理フローを示すフローチャート。

【図 24】遅延時間対応制御機能 c 及び共有メモリに設けられる情報を示す概略構成図。

【図 25】再接続開始時間調節機構 c の処理フローを示すフローチャート。

【図 26】優先待ち状態解除機構 c の処理フローを示すフローチャート。

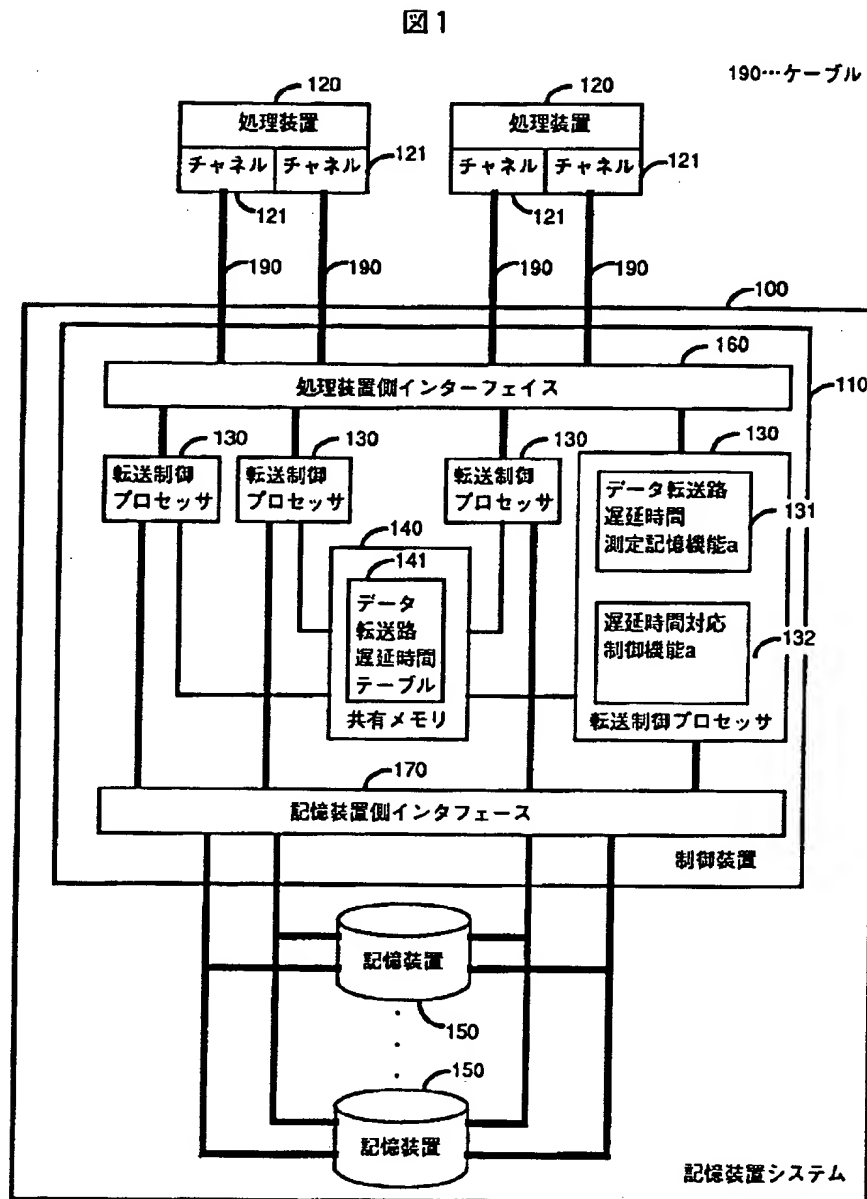
【符号の説明】

1 0 0 ……記憶装置システム、1 1 0 ……制御装置、1 2 0 ……処理装置、1 3 0 ……転送制御プロセッサ、1 3 1 ……データ転送路遅延時間測定記憶機能 a、1 3 2 ……遅延時間対応制御機能 a、1 4 0 ……共有メモリ、1 4 1 ……データ転送路遅延時間テーブル、1 5 0 ……記憶装

置、2000…データ転送路、1300…バスグループ
 対応遅延時間統計テーブル、1310…データ転送路
 遅延時間測定記憶機能、1320…遅延時間対応制御

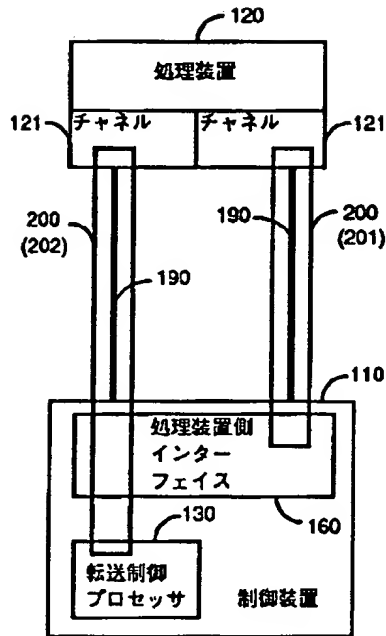
機能b、2000…処理装置対応遅延時間統計テー
 ブル、2010…データ転送路遅延時間測定記憶機能
 c、2020…遅延時間対応制御機能c。

【図1】



【図2】

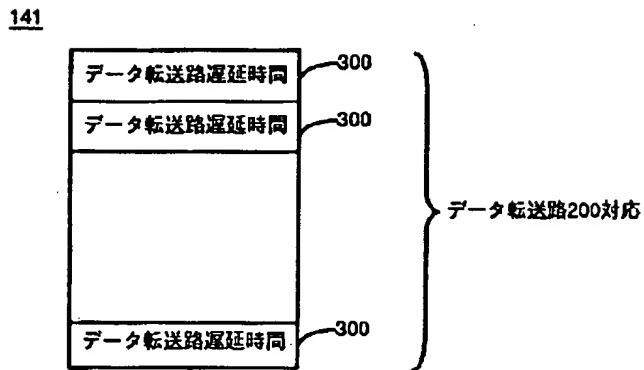
図2



190...ケーブル
200...データ転送路
201...データ転送路a
202...データ転送路b

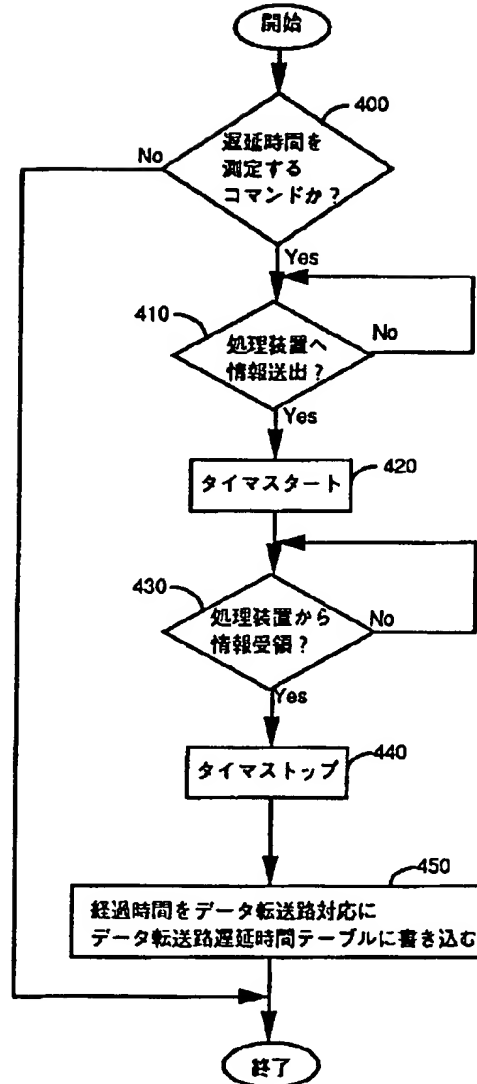
【図3】

図3

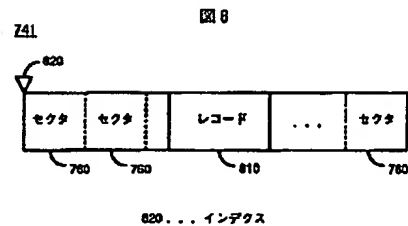


【図4】

図4

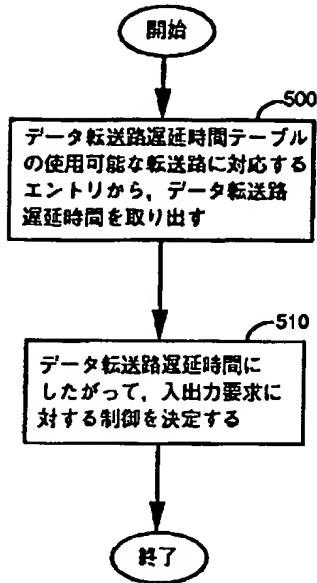


【図8】



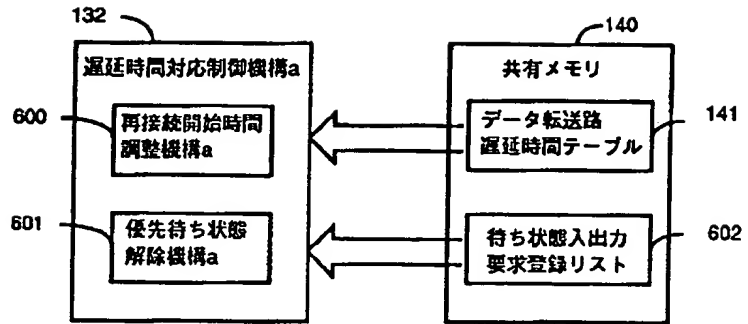
【図5】

図5



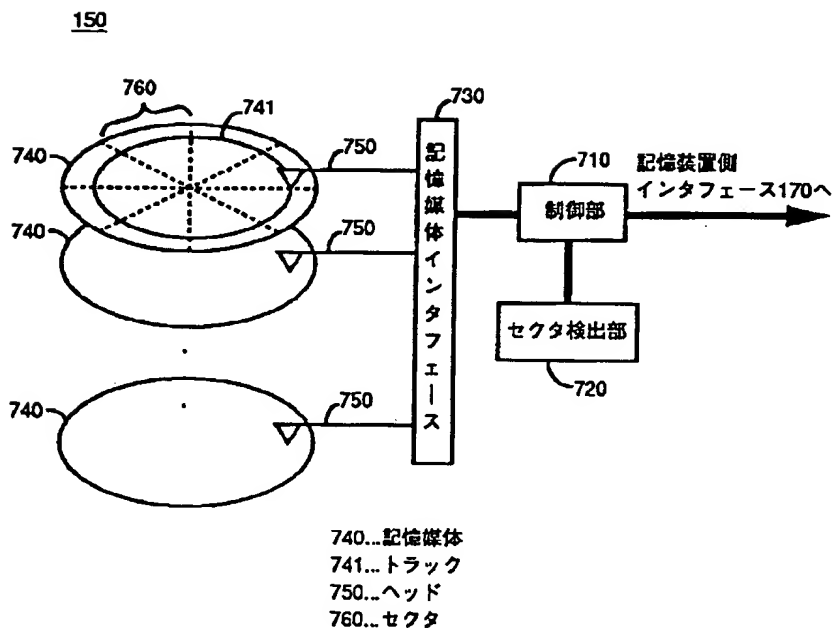
【図6】

図6



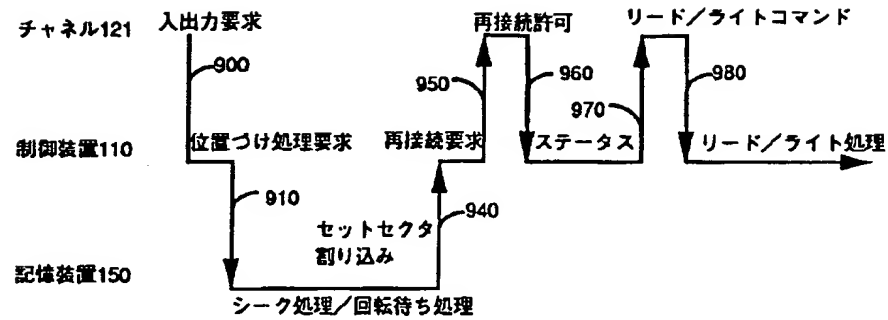
【図7】

図7



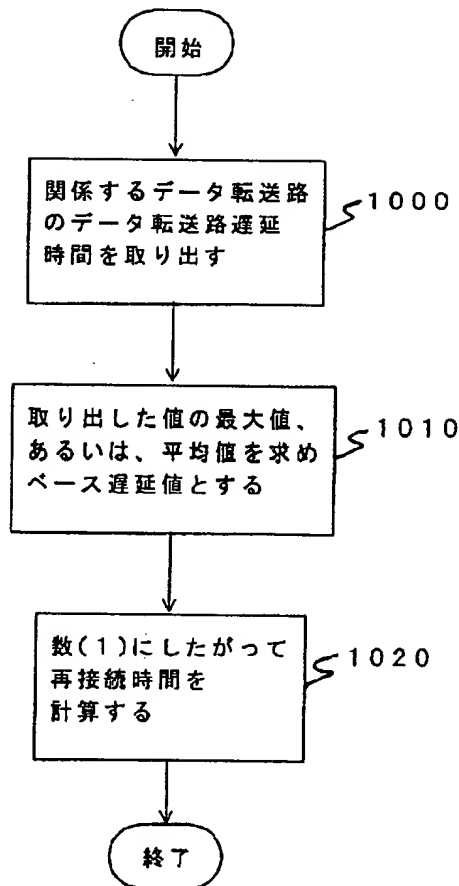
【図9】

図9



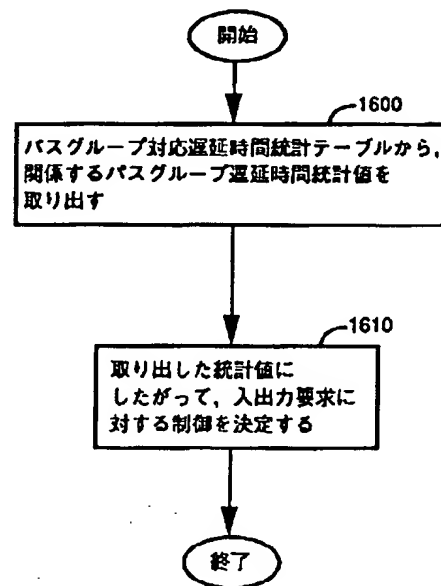
【図10】

図10



【図16】

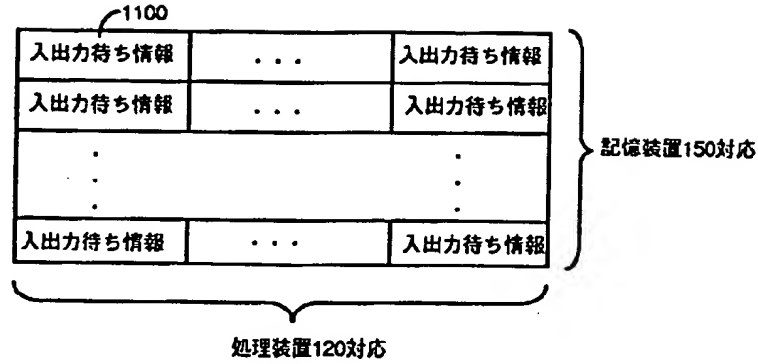
図16



【図11】

図11

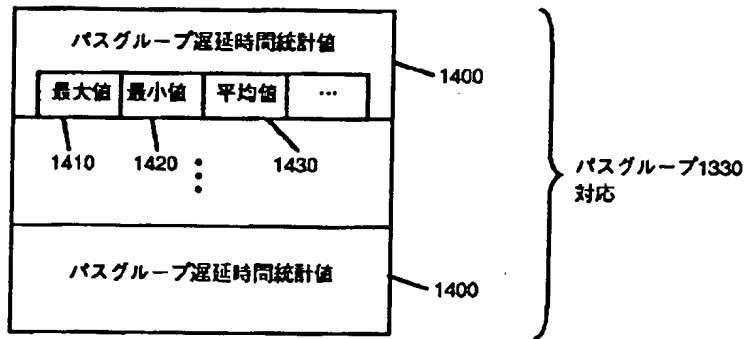
602



【図14】

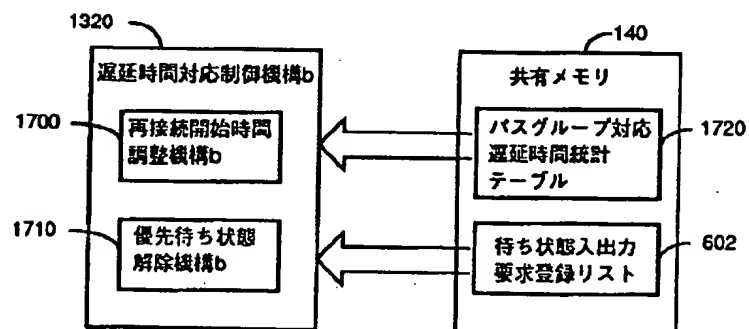
図14

1300



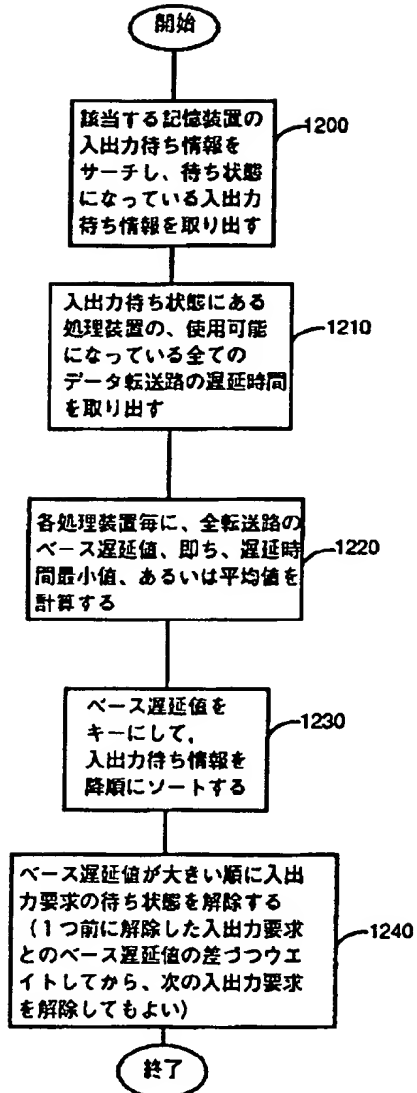
【図17】

図17



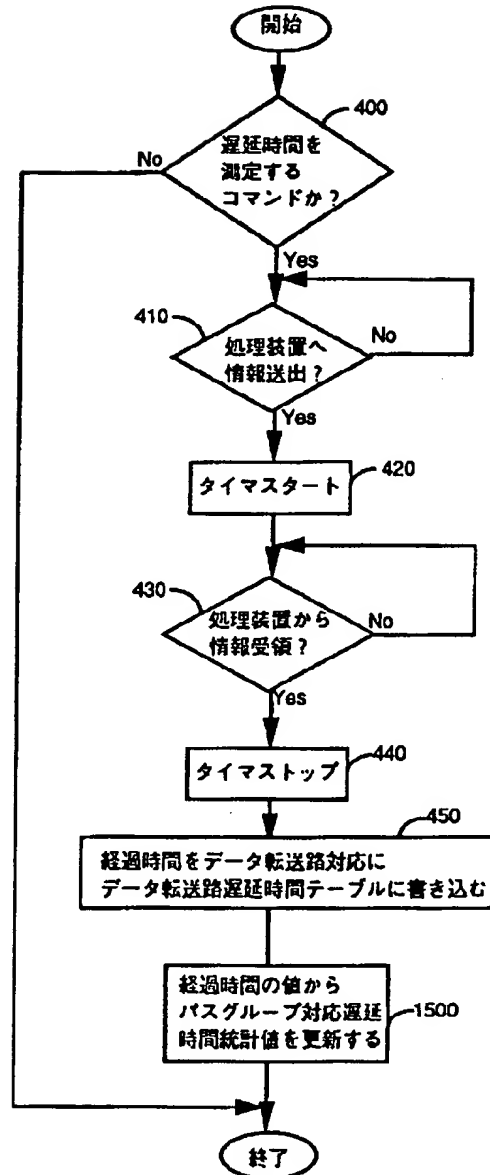
【図12】

図12



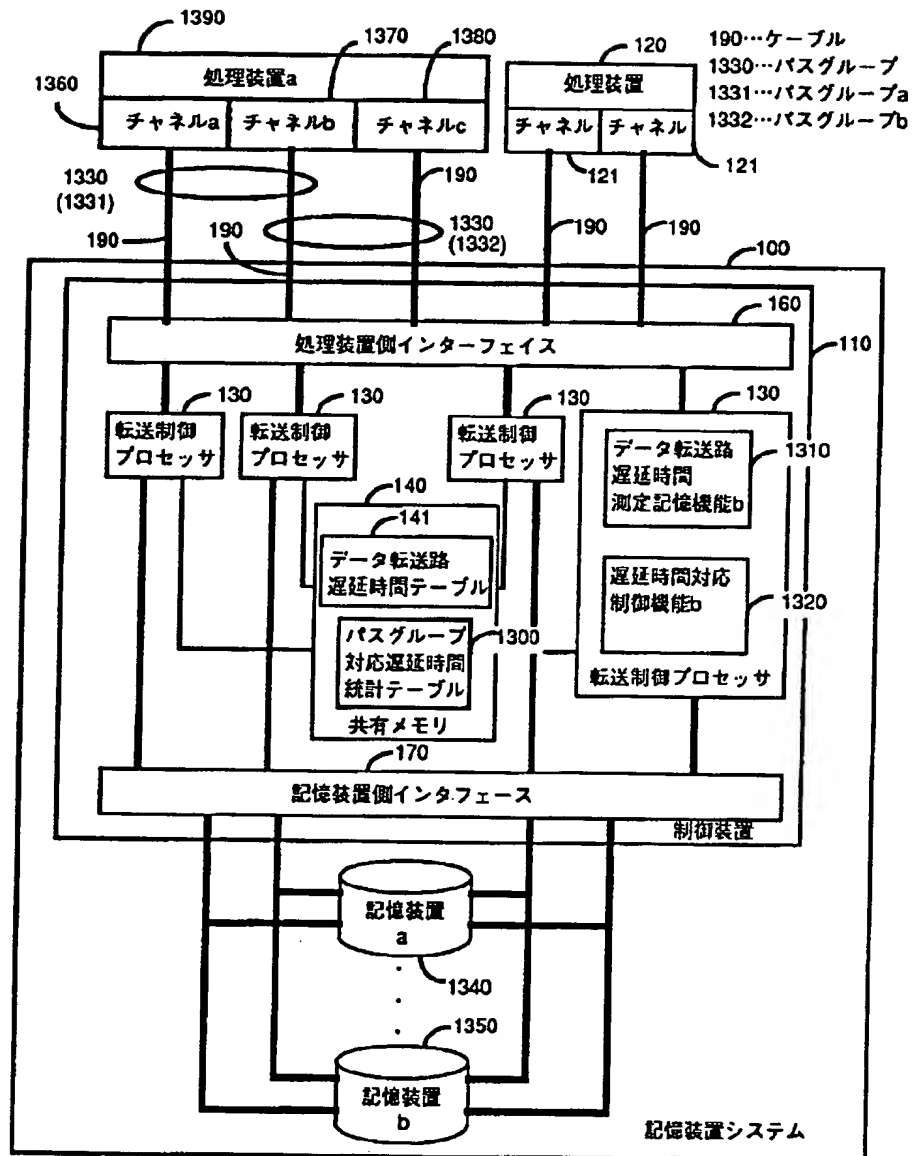
【図15】

図15



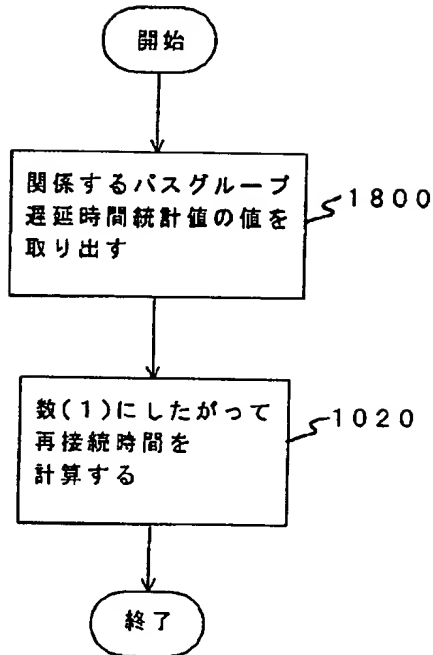
【図13】

図13



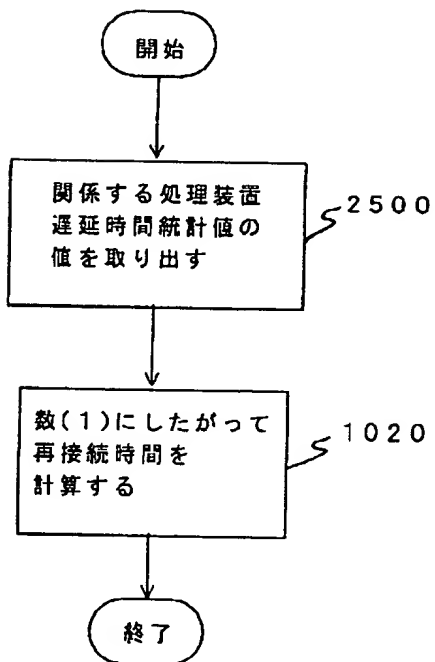
【図18】

図18



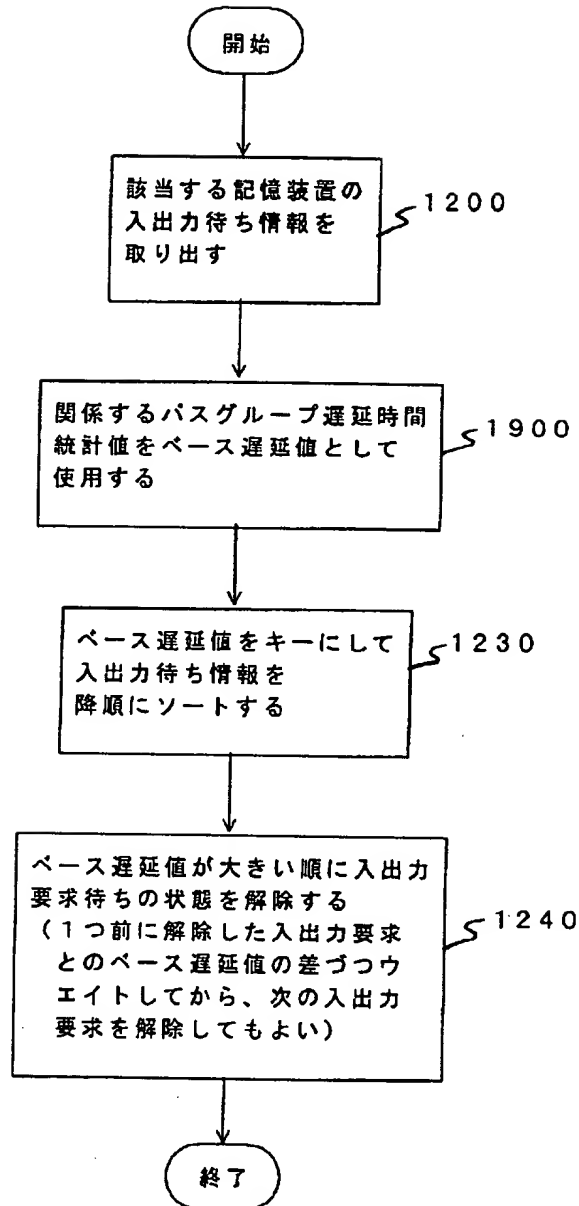
【図25】

図25



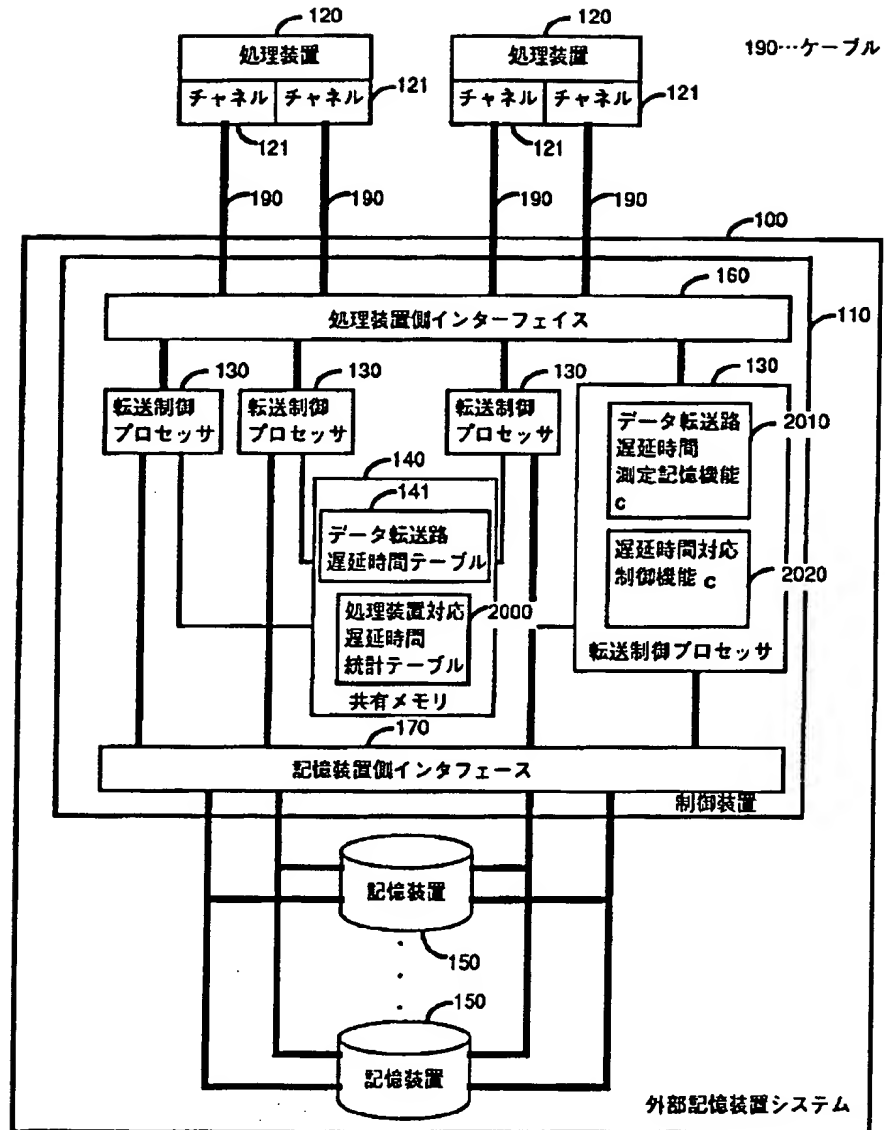
【図19】

図19



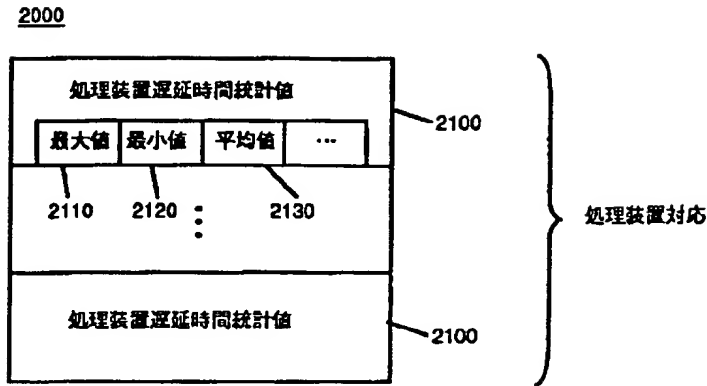
【図20】

図20



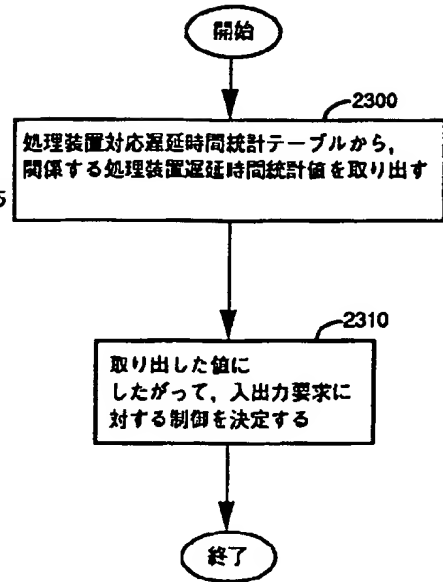
【図21】

図21



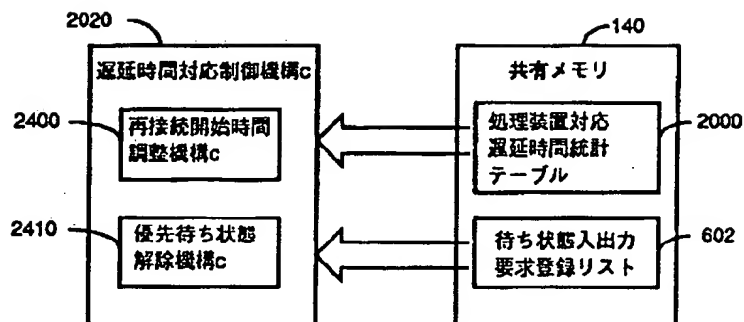
【図23】

図23



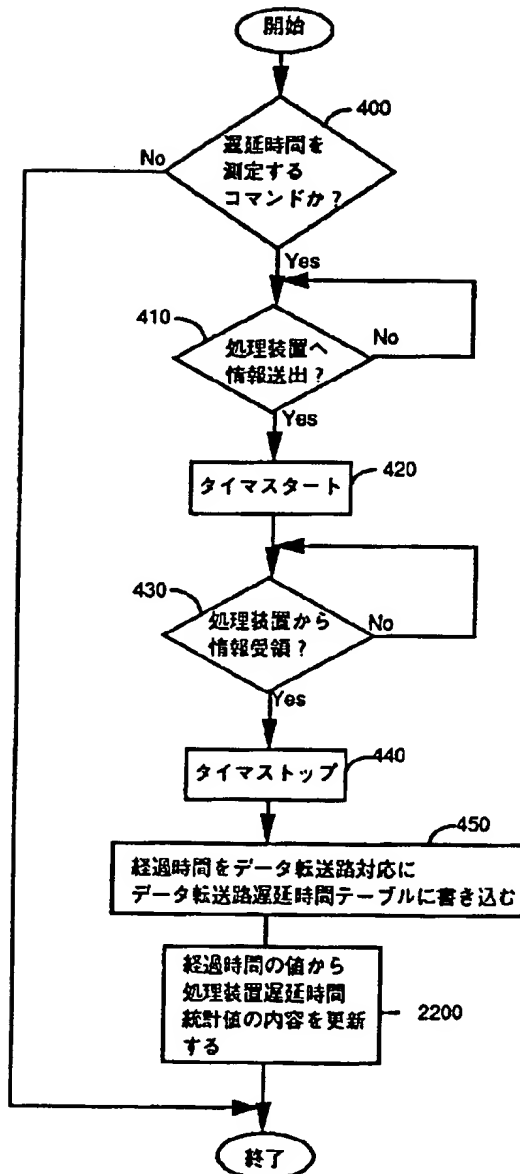
【図24】

図24



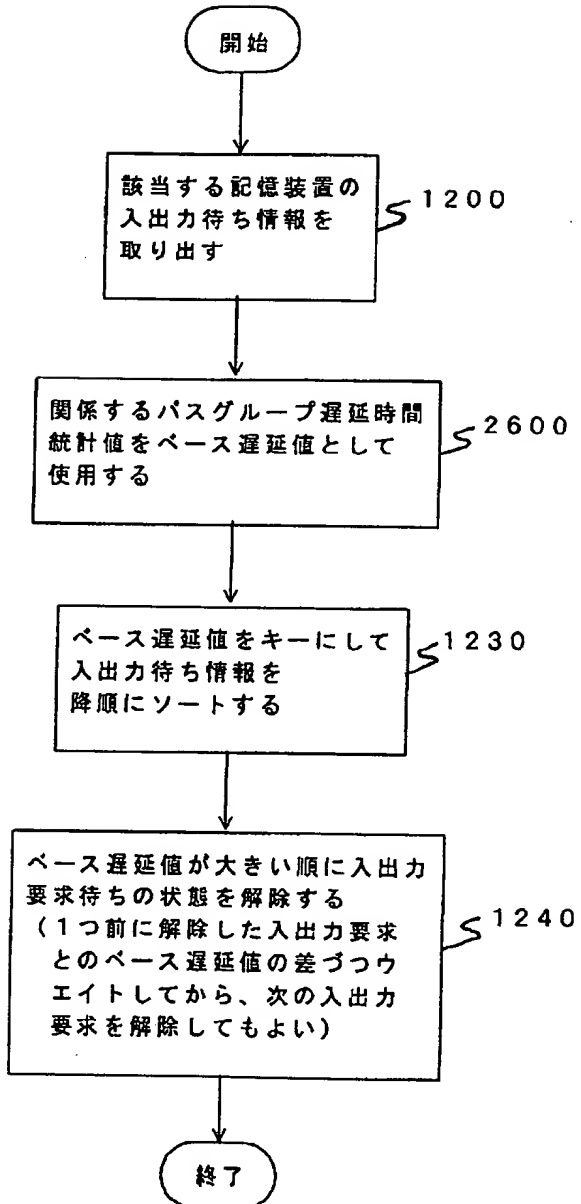
【図22】

図22



【図26】

図26



【手続補正書】

【提出日】平成6年3月22日

【手続補正1】

【補正対象書類名】明細書

【補正対象項目名】図面の簡単な説明

【補正方法】変更

【補正内容】

【図面の簡単な説明】

【図1】本発明の第1の実施例の計算機システムの構成図。

【図2】データ転送路の定義を説明するための概念図。

【図3】データ転送路遅延時間テーブル141の構成図。

【図4】データ転送路遅延時間測定記憶機能aの処理フローを示すフローチャート。

【図5】遅延時間対応制御機能aの処理フローを示すフローチャート。

【図6】遅延時間対応制御機能a及び共有メモリに設けられる情報を示す概略構成図。

【図7】記憶装置の構成を示す概略構成図。

【図8】記憶媒体上のトラックの構成図。

【図9】レコードの一般的なアクセス例を示すタイミングチャート。

【図10】再接続開始時間調節機構aの処理フローを示すフローチャート。

【図11】待ち状態入出力要求登録テーブルの構成図。

【図12】優先待ち状態解除機構aの処理フローを示すフローチャート。

【図13】本発明の第2の実施例の計算機システムの構成図。

【図14】バスグループ対応遅延時間統計テーブルの構成図。

【図15】データ転送路遅延時間測定記憶機能bの処理フローを示すフローチャート。

【図16】遅延時間対応制御機能bの処理フローを示すフローチャート。

【図17】遅延時間対応制御機能b及び共有メモリに設けられる情報を示す概略構成図。

【図18】再接続開始時間調節機構bの処理フローを示すフローチャート。

【図19】優先待ち状態解除機構bの処理フローを示すフローチャート。

【図20】本発明の第3の実施例の計算機システムの構成図。

【図21】処理装置対応遅延時間統計テーブルの構成図。

【図22】データ転送路遅延時間測定記憶機能cの処理フローを示すフローチャート。

【図23】遅延時間対応制御機能cの処理フローを示すフローチャート。

【図24】遅延時間対応制御機能c及び共有メモリに設けられる情報を示す概略構成図。

【図25】再接続開始時間調節機構cの処理フローを示すフローチャート。

【図26】優先待ち状態解除機構cの処理フローを示すフローチャート。

【符号の説明】

100…記憶装置システム、110…制御装置、120…処理装置、130…転送制御プロセッサ、131…データ転送路遅延時間測定記憶機能a、132…遅延時間対応制御機能a、140…共有メモリ、141…データ転送路遅延時間テーブル、150…記憶装置、200…データ転送路、1300…バスグループ対応遅延時間統計テーブル、1310…データ転送路遅延時間測定記憶機能、1320…遅延時間対応制御機能b、2000…処理装置対応遅延時間統計テーブル、2010…データ転送路遅延時間測定記憶機能c、2020…遅延時間対応制御機能c。

フロントページの続き

(72)発明者 中村 勝憲

神奈川県小田原市国府津2880番地 株式会社日立製作所ストレージシステム事業部内

(72)発明者 白柳 芳朗

神奈川県小田原市国府津2880番地 株式会社日立製作所ストレージシステム事業部内

(72)発明者 山本 彰

神奈川県川崎市麻生区王禅寺1099番地 株式会社日立製作所システム開発研究所内